Title
A chaotic approach to rainfall disaggregation

Journal
Water Resources Research, 37(1)

ISSN
0043-1397

Authors
Sivakumar, B
Sorooshian, S
Vijai Gupta, H
et al.

Publication Date
2001

DOI
10.1029/2000WR900196

WATER RESOURCES RESEARCH, VOL. 37, NO. 1, PAGES 61–72, JANUARY 2001

A chaotic approach to rainfall disaggregation

Bellie Sivakumar,[1] Soroosh Sorooshian, Hoshin Vijai Gupta, and Xiaogang Gao
Department of Hydrology and Water Resources, University of Arizona, Tucson

Abstract. The importance of high-resolution rainfall data to understanding the intricacies of the dynamics of hydrological processes and describing them in a sophisticated and accurate way has been increasingly realized. The last decade has witnessed a number of studies and numerous approaches to the possibility of transformation of rainfall data from one scale to another, nearly unanimously pointing to such a possibility. However, an important limitation of such approaches is that they treat the rainfall process as a realization of a stochastic process, and therefore there seems to be a lack of connection between the structure of the models and the underlying physics of the rainfall process. The present study introduces a new framework based on the notion of deterministic chaos to investigate the behavior of the dynamics of rainfall transformation between different temporal scales aimed toward establishing this connection. Rainfall data of successively doubled resolutions (i.e., 6, 12, 24, 48, 96, and 192 hours) observed at Leaf River basin, in the state of Mississippi, United States of America, are studied. The correlation dimension method is employed to investigate the presence of chaos in the rainfall transformation. The finite and low correlation dimensions obtained for the distributions of weights between rainfall data of different scales indicate the existence of chaos in the rainfall transformation, suggesting the applicability of a chaotic model. The formulation of a simple chaotic disaggregation model and its application to the Leaf River rainfall data provides encouraging results with practical potential. The disaggregation model results themselves indicate the presence of chaos in the dynamics of rainfall transformation, providing support for the results obtained using the correlation dimension method.

1. Introduction

The lack of high-resolution temporal and spatial rainfall data has been one of the most prominent limiting factors in hydrological, meteorological, and agricultural calculations. The recent shift of our focus to deal with complex problems, such as pollution transport, rainfall-related pollution effects on treatment plants, runoff-induced washoff from impermeable surfaces, soil water infiltration movement, and water erosion, only indicates the added uncertainties on the outcomes if the required quality of data is not available. One possible way to solve this problem is to collect high-resolution data relevant for the problem in question. However, this is costly and time-consuming and therefore hardly a competitive alternative in practice. As a result, the only alternative seems to be to try to transform the available data from one time and space scale to another.

The last decade has witnessed a number of studies investigating the possibility of transformation of rainfall data from one scale to another [e.g., *Hershenhorn and Woolhiser*, 1987; *Arnold and Williams*, 1989; *Econopouly et al.*, 1990; *Bo et al.*, 1994; *Glasbey et al.*, 1995; *Perica and Foufoula-Georgiou*, 1996; *Connolly et al.*, 1998; *Olsson*, 1998; *Olsson and Berndtsson*, 1998]. *Hershenhorn and Woolhiser* [1987] developed a stochastic model to disaggregate daily rainfall into a number of individual storms in a day. Each storm's starting time, duration, and amount were simulated using rainfall on the day and on the preceding and following days. However, the generation of internal storm structure was not addressed in their study. The model was somewhat modified by *Econopouly et al.* [1990], who also investigated the spatial transferability of the model parameters. *Arnold and Williams* [1989] proposed a simple stochastic model to generate half-hourly rainfall intensity from daily rainfall. The model assumed that the daily rainfall fell in only one event. *Cowpertwait et al.* [1996] developed a stochastic disaggregation model for rainfall observed in the United Kingdom, which allowed historical or generated hourly data to be disaggregated into totals for shorter time intervals. They also observed the necessity of carrying out some smoothing of the disaggregated totals to obtain realistic storm profiles. *Connolly et al.* [1998] developed a model which allows disaggregation of daily rainfall into multiple events in a day and the simulation of time-varying intensity within each event.

Bo et al. [1994] used the Bartlett-Lewis rectangular pulses (BLRP) model developed by *Rodriguez-Iturbe et al.* [1987, 1988] to disaggregate daily rainfall into hourly values. They argued that the successful result was due to a scaling (power law) behavior of the power spectrum. Another approach was proposed by *Glasbey et al.* [1995], who modified the BLRP model to simulate hourly data consistent with observed daily totals. *Perica and Foufoula-Georgiou* [1996] performed scaling-based disaggregation for spatial rainfall. They developed a disaggregation model based on the (empirically observed) scaling of probability distributions of rainfall fluctuations and correlation between the scaling parameters and the convective available potential energy. *Olsson* [1998] employed a cascade scheme [e.g., *Gupta and Waymire*, 1993] to model the temporal small-scale structure of rainfall. The main difference between

[1]Now at Department of Land, Air and Water Resources, University of California, Davis.

Paper number 2000WR900196.
0043-1397/01/2000WR900196$09.00

this model and most other previously developed cascade models is the exact conservation of rainfall volume between successive cascade levels. This is termed a microcanonical property as opposed to canonical cascades where the volume is only on average conserved [e.g., *Mandelbrot*, 1974; *Schertzer and Lovejoy*, 1987]. Another consequence of this cascade structure is that the weights are not mutually independent but have a correlation of -1. However, the pairs of weights associated with different branchings are assumed to be mutually independent. *Olsson* [1998] claims that the model differs in some fundamental respects from the majority of random cascade models, even using a similar microcanonical approach [e.g., *Carsteanu and Foufoula-Georgiou*, 1996], because of the variation in the correlation between adjacent pairs of weights in individual rainfall events. *Olsson and Berndtsson* [1998] employed the scaling-based cascade model developed by *Olsson* [1998] to disaggregate the daily seasonal rainfall time series of 3 years observed in Sweden to 45-min resolution. The disaggregated data were shown to reproduce very well many fundamental characteristics of the observed 45-min data, for example, division between rainy and dry periods, the event structure, and the scaling behavior. The results thus demonstrated the potential of scaling-based approaches in hydrological applications involving rainfall.

It is encouraging that the studies conducted thus far on the transformation of rainfall data from one scale to another nearly unanimously support such a possibility. Particularly, the scaling-based studies that employed multifractal random cascade schemes seem to be useful to improve our understanding of the rainfall phenomenon and the possibility of rainfall disaggregation and therefore may have the prospect of tremendous practical significance. However, the main concern of such approaches is naturally the connection between the structure of the model on the one hand and the underlying physics of the rainfall process on the other. Establishing this connection remains a crucial task of research in the present field [*Olsson*, 1998]. In other words, the multifractal random cascade approaches treat the data as a realization of a stochastic process whose prevalent characteristic is the multifractal spectrum, and therefore they may not be able to account for the uniqueness of the data at hand [*Puente and Obregon*, 1996]. Therefore an approach that seeks to understand the whole and unique data set, in addition to preserving the multifractal spectrum, other important qualifiers of the records, and the overall appearance of the data, could better serve the purpose. In this regard the notion of deterministic chaos (that seemingly irregular-looking behavior can be the result of simple deterministic systems influenced by a few nonlinear interdependent variables) and the related methods of data processing seem to have a great potential.

The application of the concept of deterministic chaos to hydrological processes has recently received considerable attention. A number of studies have investigated the existence of chaos in various hydrological processes, such as rainfall [e.g., *Hense*, 1987; *Rodriguez-Iturbe et al.*, 1989; *Sharifi et al.*, 1990; *Berndtsson et al.*, 1994; *Jayawardena and Lai*, 1994; *Puente and Obregon*, 1996; *Sivakumar et al.*, 1998; 1999a, 1999b; *Sivakumar*, 1999, also Chaotic analysis of rainfall observed at different scales, submitted to *Journal of Hydrology*, 2000 (hereinafter referred to as Sivakumar, submitted manuscript, 2000], streamflow [e.g., *Jayawardena and Lai*, 1994; *Porporato and Ridolfi*, 1996, 1997], and lake volume [e.g., *Sangoyomi et al.*, 1996]. A few studies have also attempted short-term predictions [e.g., *Jayawardena and Lai*, 1994; *Porporato and Ridolfi*, 1996, 1997; *Sivakumar*, 1999; *Sivakumar et al.*, 1999a] and noise reduction [e.g., *Berndtsson et al.*, 1994; *Porporato and Ridolfi*, 1997; *Sivakumar et al.*, 1999b; *Sivakumar*, 1999] in hydrological processes. The outcomes of such studies are, indeed, encouraging, as almost all of them have provided evidence on the existence of chaos in hydrological processes, thus revealing that the notion of deterministic chaos could be a proper alternative to the stochastic framework.

In spite of its recent widespread application in various hydrological processes, to the authors' knowledge, the suitability of the concept of chaos for transformation of hydrological data from one scale to another has not been studied so far and therefore provides the driving force for the present study. The objectives of the present study are twofold: (1) to investigate whether or not the transformation of rainfall data from one scale to another follows chaotic behavior and (2) to formulate a simple chaotic model for transformation of rainfall data from one scale to another. This study focuses only on the transformation of rainfall data from one timescale (low resolution) to another (high resolution), that is, temporal transformation or disaggregation. Rainfall data of different resolutions, that is, 6, 12, 24, 48, 96, and 192 hours, over a period of about 25 years observed at Leaf River basin, in the state of Mississippi, United States of America, are considered for the investigation. The correlation dimension method is employed to investigate the presence (or absence) of chaos.

The organization of this paper is as follows. In section 2 the correlation dimension method is reviewed. The method is then employed to investigate whether or not the transformation of rainfall data from one scale to another exhibits chaotic behavior. A chaotic model is formulated in section 3 for transforming rainfall data from one scale to another. The effectiveness of the model for rainfall disaggregation is tested on data observed at Leaf River basin. The model also describes an inverse approach to identifying chaos using the results of disaggregation. These results lead to a general discussion in section 4 concerning the question of whether a hypothesis of deterministic chaos is valid for transformation of rainfall data from one scale to another.

2. Investigation of Chaos in Rainfall Transformation

2.1. Introduction

The science of chaos is a burgeoning field, and the available methods to investigate the existence of chaos in a time series are still in a state of infancy. However, the considerable attention that the theory has received in almost all fields of natural and physical sciences has motivated improvements in existing methods for the diagnosis of chaos and the proposal of new ones. The methods available thus far are the correlation dimension method [e.g., *Grassberger and Procaccia*, 1983a, 1983b], the nonlinear prediction method [e.g., *Farmer and Sidorowich*, 1987; *Casdagli*, 1989; *Sugihara and May*, 1990] including deterministic versus stochastic diagram [e.g., *Casdagli*, 1991], the Lyapunov exponent method [e.g., *Wolf et al.*, 1985], the Kolmogorov entropy method [e.g., *Grassberger and Procaccia*, 1983c], the surrogate data method [e.g., *Theiler et al.*, 1992], and the linear and nonlinear redundancies [e.g., *Palus*, 1995; *Prichard and Theiler*, 1995]. Among these the correlation dimension method has been the most widely used one for the investigation of deterministic chaos in hydrological phenomena [e.g., *Hense*, 1987; *Rodriguez-Iturbe et al.*, 1989; *Sharifi et al.*, 1990; *Berndtsson et al.*, 1994; *Jayawardena and Lai*, 1994; *Puente and Obregon*, 1996; *Sangoyomi et al.*, 1996; *Porporato and Ridolfi*, 1996, 1997; *Sivakumar et al.*, 1998, 1999a; *Sivakumar*, 1999, submitted manuscript, 2000]. In the present study, the correlation dimension method is employed, and the presence of a low-dimensional attractor (a geometric object which characterizes the long-term behavior of a system in the phase space) is taken as an indication of chaos.

It is relevant to note that the application of chaos identification methods, particularly the correlation dimension method, to hydrological time series and the reported results have very often been questioned because of the fundamental assumptions with which the methods have been developed, that is, that the time series is infinite and noise-free. Important issues in the application of chaos identification methods to hydrological data, for example, data size, noise, delay time, etc., and the validity of chaos theory in hydrology have been discussed in detail by *Sivakumar* [2000] and therefore are not reported herein. It is relevant to note, however, that the studies by *Sivakumar* [1999, 2000] reveal that the presence of noise in the data does not significantly influence the correlation dimension estimates (though it significantly influences the prediction accuracy estimates). This suggests that the correlation dimension may be used as a preliminary indicator to identify the existence of chaos in the transformation of (noisy) rainfall data from one scale to another, before attempting detailed analysis such as application of noise reduction procedures. In this context, no attempt is made, in the present study, to reduce the noise present in the rainfall time series, and the raw data as recorded are analyzed. However, it is believed that the estimation of the correlation dimension requires a large number of data points. Such a belief is based on the assumption that the data size depends on the embedding dimension in the phase-space reconstruction. However, the study by *Sivakumar* [2000] reveals that the data size depends on the type and dimension of the underlying process, rather than the embedding dimension, and therefore the problem of data size is not as severe as it is believed to be (see *Sivakumar* [2000] for details). It is the authors' belief that rainfall time series observed over a period of 25 years is reasonably sufficient to understand the dynamics of rainfall transformation.

2.2. Correlation Dimension Method

The goal of determining the dimension of an attractor is that the dimensionality of an attractor furnishes information on the number of dominant variables present in the evolution of the corresponding dynamical system. Dimension analysis will also reveal the extent to which the variations in the time series are concentrated on a subset of the space of all possible variations. The central idea behind the application of the dimension approach is that systems whose dynamics are governed by stochastic processes are thought to have an infinite value for the dimension. A finite, noninteger value of the dimension is considered to be an indication of the presence of chaos.

Correlation dimension is a measure of the extent to which the presence of a data point affects the position of the other points lying on the attractor. The correlation dimension method uses the correlation integral (or function) for distinguishing between chaotic and stochastic behaviors. The concept of the correlation integral is that an irregular-looking process arising from deterministic dynamics will have a limited number of degrees of freedom equal to the smallest number of first-order differential equations that capture the most important features of the dynamics. Thus, when one constructs phase spaces of increasing dimension for an infinite data set, a point will be reached where the dimension equals the number of degrees of freedom and beyond which increasing the dimension of the representation will not have any significant effect on the correlation dimension.

Although many algorithms have been formulated for the computation of the correlation dimension of a time series, the Grassberger-Procaccia algorithm [*Grassberger and Procaccia*, 1983a, 1983b] has received a lot of attention and has been applied in a number of studies [e.g., *Hense*, 1987; *Rodriguez-Iturbe et al.*, 1989; *Sharifi et al.*, 1990; *Berndtsson et al.*, 1994; *Jayawardena and Lai*, 1994; *Puente and Obregon*, 1996; *Sangoyomi et al.*, 1996; *Porporato and Ridolfi*, 1996, 1997; *Sivakumar et al.*, 1998, 1999a; *Sivakumar*, 1999, submitted manuscript, 2000] and therefore is employed in the present study too. The algorithm uses the phase-space reconstruction of the time series. For a scalar time series X_i, where $i = 1, 2, \ldots, N$, the phase space can be reconstructed using the method of delays [e.g., *Takens*, 1980]. The basic idea in the method of delays is that the evolution of any single variable of a system is determined by the other variables with which it interacts. Information about the relevant variables is thus implicitly contained in the history of any single variable. On the basis of this an "equivalent" phase space can be reconstructed by assigning an element of the time series X_i and its successive delays as coordinates of a new vector time series

$$\mathbf{Y}_j = (X_j, X_{j+\tau}, X_{j+2\tau}, \ldots, X_{j+(m-1)\tau}), \quad (1)$$

where $j = 1, 2, \ldots, N - (m - 1)\tau/\Delta t$, m is the dimension of the vector $\mathbf{Y}_j$, also called the embedding dimension, and τ is a delay time taken to be some suitable multiple of the sampling time Δt [*Packard et al.*, 1980; *Takens*, 1980]. According to the embedding theorem of *Takens* [1980], to characterize a dynamic system with an attractor dimension d, an $(m = 2d + 1)$-dimensional phase space is required. However, *Abarbanel et al.* [1990] suggested that $m > d$ would be sufficient. For an m-dimensional phase space the correlation function $C(r)$ is given by

$$C(r) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{\substack{i,j \\ (1\le i<j\le N)}} H(r - |Y_i - Y_j|), \quad (2)$$

where H is the Heaviside step function, with $H(u) = 1$ for $u > 0$ and $H(u) = 0$ for $u \le 0$, where $u = r - |\mathbf{Y}_i - \mathbf{Y}_j|$, r is the radius of sphere centered on $\mathbf{Y}_i$ or $\mathbf{Y}_j$, and N is the number of data points.

If the time series is characterized by an attractor, then for positive values of r the correlation function $C(r)$ is related to the radius r by the following relation:

$$\underset{\substack{r\to 0 \\ N\to\infty}}{C(r)} \sim \alpha r^{\nu}, \quad (3)$$

where α is constant and ν is the correlation exponent or the slope of the log $C(r)$ versus log r plot given by

$$\nu = \lim_{\substack{r\to 0 \\ N\to\infty}} \frac{\log C(r)}{\log r}. \quad (4)$$

The slope is generally estimated by a least squares fit of a straight line over a certain range of r, called the scaling region.

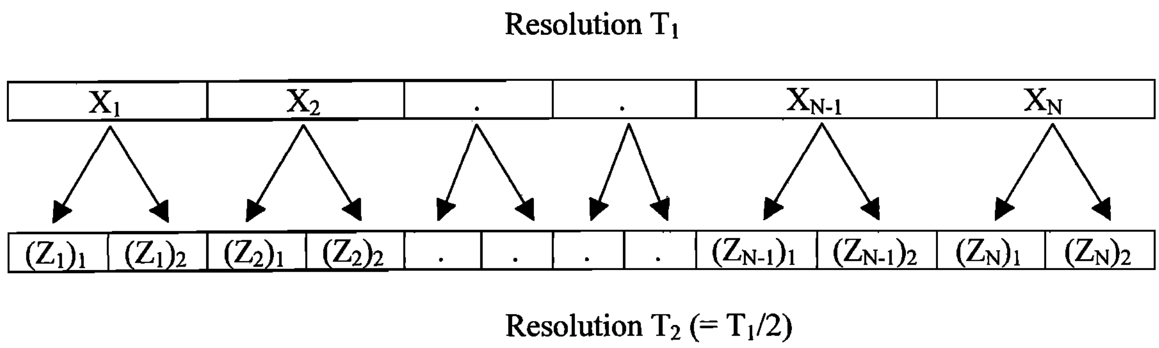

Figure 1. Schematic representation of distributions of weights of rainfall transformation.

For a finite data set, such as the rainfall data series, it is clear that there is a separation r below which there are no pairs of points; that is, it is "depopulated." At the other extreme, when the value of r exceeds the set diameter, the correlation function increases no further; that is, it is "saturated." Therefore, for a finite data set, the region sandwiched between the depopulation region and the saturation region is considered as the scaling region. A somewhat better way to identify the scaling region is to estimate the local slope given by $d[\log C(r)]/d[\log r]$.

To observe whether chaos exists, the correlation exponent (or local slope) values are plotted against the corresponding embedding dimension values. If the value of the correlation exponent is finite, low, and noninteger, then the system is considered to exhibit low-dimensional chaos. The saturation value of the correlation exponent is defined as the correlation dimension of the attractor. The nearest integer above the saturation value provides the minimum number of phase spaces or variables necessary to model the dynamics of the attractor. On the contrary, if the correlation exponent increases without bound with increase in the embedding dimension, then the system under investigation is considered as stochastic [*Fraedrich*, 1986].

2.3. Analysis of Rainfall Data

2.3.1. Introduction. When the purpose is to investigate the existence of chaos in a hydrological time series of only a particular resolution, such as daily or hourly, then the task is simple, as the correlation dimension method can directly be employed to the time series. However, such a straightforward approach may not be possible when the task at hand is to investigate the existence of chaos in the dynamics of transformation of data from one scale to another, and therefore the problem becomes more complicated. One way to solve this problem may be to analyze time series of different resolutions so as to understand their behaviors (chaotic or stochastic). However, while such an approach will no doubt provide indications regarding the behavior of time series at the different resolutions analyzed, the question remains whether this can reveal anything about the behavior of the transformations between them. For example, if time series of different resolutions, such as daily and hourly, exhibit chaotic (or stochastic) behavior, this does not mean that the transformations are also chaotic (or stochastic). In other words, it may not be possible to relate the behavior of individual time series of different resolutions to that of the overall transformations between them. Another possible way to solve this problem is to determine how a time series of a particular resolution is transformed (aggregated/disaggregated) into that of another resolution using the distributions of weights between the data resolutions. The consideration of distributions of weights between data of different resolutions seems to be a much more promising approach than the former one to obtain important information regarding the behavior of transformation of time series from one scale to another. This approach therefore is employed in the present study to investigate the behavior of transformation of rainfall data from one scale to another.

2.3.2. Distributions of weights. Let us assume that we have a rainfall time series X_i, $i = 1, 2, \ldots, N$, at a certain resolution T_1, and the task at hand is to obtain the rainfall values $(Z_i)_k$, $k = 1, 2, \ldots, p$, at a higher resolution T_2, where $p = T_1/T_2$. Let us also assume that the values of X_i are distributed into $(Z_i)_k$ according to $(Z_i)_k = (W_i)_k X_i$, where $(W_i)_k$ are the weights of distributions of X_i to $(Z_i)_k$ and $\Sigma_{k=1}^{p} (W_i)_k = 1$. This indicates that the transformation of rainfall data from one scale to another is possible only if the weights of distributions of data between the two scales are available. This is why the determination of the behavior (chaotic or stochastic) of distributions of weights is essential in understanding the behavior of transformation of data from one scale to another. In practice, however, the weights are not known a priori and therefore have to be computed from the values of X_i and $(Z_i)_k$ in the model formulation stage. In the present study, for convenience, only rainfall values at successively doubled temporal resolutions are considered to compute the weights and in the subsequent development of the model. This means that $p = T_1/T_2 = 2$. A schematic picture depicting the approach employed in the present study is shown in Figure 1.

2.3.3. Study area and data used. In the present study, rainfall data observed at Leaf River basin in the state of Mississippi, United States of America, is used. The basin is located between 31°42′25″ latitude and 89°24′25″ longitude in the southern region of the state of Mississippi. The climate in this region is humid subtropical, characterized by short, temperate winters, long, hot summers, and rainfall that is fairly evenly distributed throughout the year. However, the region is subject to periods of both drought and flood, and the climate rarely seems to bring "average" conditions. Prevailing southerly winds provide moisture of high humidities and potential discomfort from May through September. Locally violent and destructive thunderstorms are a threat on an average of about 60 days a year. Hurricanes and tornadoes are a particular danger, especially during the spring season.

The mean annual temperature is about 18.3°C, the minimum mean of 5.6°C occurring in January and the maximum mean of

Table 1. Characteristics of Rainfall Data of Different Resolutions

Parameter	192 Hours	96 Hours	48 Hours	24 Hours	12 Hours	6 Hours
Number of data	1,024	2,048	4,096	8,192	16,384	32,768
Mean	32.24	16.12	8.06	4.03	2.02	1.01
Standard deviation	31.90	22.08	15.61	10.47	6.71	4.12
Variance	1,017.62	487.62	243.56	109.46	44.97	16.96
Maximum value	234.03	221.52	221.52	221.52	116.16	87.37
Minimum value	0.00	0.00	0.00	0.00	0.00	0.00
Number of zeros	62 (6.1%)	412 (20.1%)	1,633 (39.9%)	4,467 (54.5%)	11,257 (68.7%)	24,889 (75.9%)

27.2°C occurring in July. Low temperatures drop to 8.9°C below zero, while high temperatures exceed 32°C. The mean annual precipitation is about 1350 mm and is well distributed throughout the year. March is the wettest month with a mean rainfall of about 160 mm, while October is the driest month with a mean rainfall of about 80 mm. Snowfall is less than 63 mm per year. In essence, the region has a climate characterized by the absence of severe cold in winter but by the presence of extreme heat in summer. Cold spells are usually of short duration, and the growing season is long. Rainfall is plentiful, but so are dry spells and sunshine.

For the present investigation, rainfall data observed over a period of about 25 years (January 1963 to December 1987) are used. The highest-resolution data available are at 6-hour intervals, and therefore 6-hour, 12-hour, 24-hour, 48-hour, 96-hour, and 192-hour data are used in the analysis. The weights considered for the analysis are those obtained from the transformation of only the nonzero rainfall values at a particular resolution (e.g., 12 hours) to a successively doubled higher resolution (e.g., 6 hours). The zero values are eliminated because their presence could significantly influence, as explained below, the outcomes of the correlation dimension analysis and the subsequent disaggregation model: (1) The presence of a large number of zero values may bias (lower) the correlation dimension estimate, since the reconstructed hypersurface in phase space would tend to a point when a certain value (in this case zero) occurs repeatedly in a time series [e.g., *Tsonis et al.*, 1993]. The correlation dimension results obtained by a recent study by Sivakumar (submitted manuscript, 2000) for the Leaf River rainfall support this point, as the daily rainfall series yields a lower dimension than the 2-day, 4-day, and 8-day rainfall series. For example, the daily rainfall time series in Leaf River, consisting of about 54.53% of zero values, yields a correlation dimension of about 4.82, which is significantly less than the dimension of the 8-day rainfall time series, consisting of only about 6.05% of zeros. This means that the inclusion of zero values, yielding an underestimation of the dimension, could indicate the presence of chaos even when it is actually absent. Such an outcome would certainly be misleading when the task at hand is basically to investigate whether or not chaos exists in the time series. (2) The presence of a zero value in rainfall time series of a particular resolution does not contribute anything to its disaggregation to another (higher) resolution, as it is obvious that the disaggregation values of a zero rainfall value are also zeros. (Also, the computation of weight ratios, which are input for the disaggregation model, of a zero value is not possible, as zero divided by zero is undefined.) (3) With respect to point 2 above, the presence of zero values in the rainfall time series may potentially lead to wrong outcomes also in the disaggregation procedure. For instance, if the zero values are included in the phase-space reconstruction and if the last element of the nearest neighbor to the vector containing the value to be disaggregated is zero, then the rainfall value cannot be disaggregated, as the weights are undefined (see section 3 for details about the disaggregation procedure).

Having said the above, it must also be noted that the elimination of zero rainfall values could also lead to unrealistic estimates in the correlation dimension estimation (or prediction) of the rainfall time series (or any other hydrological time series for that matter). This is because, in such estimates, the zero rainfall values are also indicative of, and equally important to understand, how the dynamics of the system evolved. Therefore it is necessary to understand the potential influences of the inclusion or exclusion of the zero values with respect to the particular task at hand and to obtain the appropriate time series for analysis. With the above knowledge of the potential positive and negative effects of the exclusion (or inclusion) of the zero values on the correlation dimension estimate, it is believed that their exclusion would have less effect on the outcomes of the dimension analysis, and therefore only nonzero values are considered here. Therefore the approach adopted in the present study (by eliminating zeros) to obtain the appropriate time series (i.e., distributions of weights) for the correlation dimension estimation is somewhat different from the one that is usually adopted in a rainfall (or any other hydrological) time series. It is relevant to note, however, that the resulting time series may still provide a time series of distributions of weights containing zero values. This is because a nonzero rainfall value at a particular resolution may be transformed to a (same) nonzero value and a zero value at some other resolution, in which case the weights will be 1 and 0, respectively. As a result, there is still a possibility for underestimation of the correlation dimension of the weights, but the underestimation, if any, will not be as severe as when the zero values are also included to obtain the weights. Also, the presence of zero values in the weights does not have any ill effects in the disaggregation procedure, as they are not used in the phase-space reconstruction but only helps one to obtain realistic estimates of the distribution of weights (1 and 0 or 0 and 1) and the disaggregation values.

2.3.4. Analysis, results, and discussions. Table 1 presents some of the important statistics of the rainfall data of six different resolutions, that is, 6, 12, 24, 48, 96, and 192 hours, observed at Leaf River basin. The corresponding statistics of the distributions of weights of rainfall data between the above successively doubled resolutions are presented in Table 2. As mentioned previously, the number of data presented in Table 2 for the distributions of weights between any two successively doubled resolutions, for example, 6 hours and 12 hours, is obtained by considering only the nonzero rainfall values at the

Table 2. Characteristics of Weights Between Different Rainfall Resolutions

Parameter	192-hour to 96-hour Resolution	96-hour to 48-hour Resolution	48-hour to 24-hour Resolution	24-hour to 12-hour Resolution	12-hour to 6-hour Resolution
Number of data	1,924	3,272	4,926	7,450	10,254
Mean	0.50	0.50	0.50	0.50	0.50
Standard deviation	0.3846	0.4315	0.4394	0.4427	0.4395
Variance	0.1479	0.1862	0.1931	0.1960	0.1932
Maximum value	1.00	1.00	1.00	1.00	1.00
Minimum value	0.00	0.00	0.00	0.00	0.00
Number of zeros	288 (14.97%)	809 (24.72%)	981 (19.91%)	1,726 (23.17%)	2,165 (21.12%)

lower resolution among the two, that is, 12 hours. A comparison of Tables 1 and 2 reveals that the number of zeros in the weights (Table 2) is significantly reduced by the exclusion of zero values in the computation of the weights. For example, the percentage of zeros in the weights between 12-hour and 6-hour resolutions is about 21% when the zeros are excluded in the computation of weights, whereas the corresponding number would be about 75% if zeros are also included. While it can be argued that the weights obtained in the former case could still lead to an underestimation of the dimension because of the presence of zeros, it is crucial to recognize that the underestimation, if any, would be significantly less compared to the latter. These observations seem to suggest that the correlation dimension values obtained in the present study for the weights could, indeed, be reasonably closer to the actual ones, if not accurate (see below for more details).

Figure 2, for example, shows the variation of the distributions of weights between rainfall data of 12-hour and 6-hour resolutions observed at Leaf River basin. A visual inspection of the time series indicates significant variations in the values. The seemingly irregular behavior of the time series does not indicate anything regarding its behavior, whether deterministic or stochastic or chaotic, and therefore additional tools are required to identify whether or not chaos exists. Similar observations are also made for the weights between rainfall data of the other successively doubled resolutions, that is, 24 and 12 hours, 48 and 24 hours, 96 and 48 hours, and 192 and 96 hours. The analyses of the distributions of weights using the correlation dimension method and the results obtained are presented below.

In the present study, the correlation functions and hence the exponents are computed, as explained in section 2.2, for distributions of weights between rainfall data observed at different successively doubled resolutions. The delay time τ is computed using the autocorrelation function method and is taken as the lag time at which the autocorrelation function first crosses the zero line [e.g., *Holzfuss and Mayer-Kress*, 1986]. A delay time value of 1 is observed for all the different sets of weights. Figure 3a shows the log $C(r)$ versus log r plots obtained with embedding dimensions from 1 to 20 for distributions of weights between rainfall data observed at 12-hour and 6-hour resolutions, while the variation of local slope against log r is shown in Figure 3b. Figures 3a and 3b indicate clear scaling regions that allow fairly accurate estimates of the correlation exponents. The relationship between the correlation exponent values and the embedding dimension values is shown in Figure 4a. It can be seen that the correlation exponent value increases with the embedding dimension up to a certain value and then saturates beyond that value. The saturation of the correlation exponent beyond a certain embedding dimension value is an indication of the existence of deterministic dynamics. The saturation value of the correlation exponent (or correlation dimension) is about 1.86. The finite and low correlation dimen-

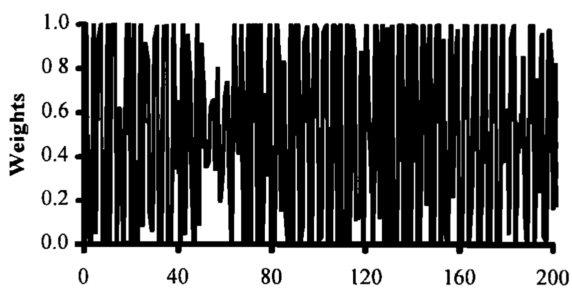

Figure 2. Variation of distributions of weights of rainfall data between 12-hour and 6-hour resolutions.

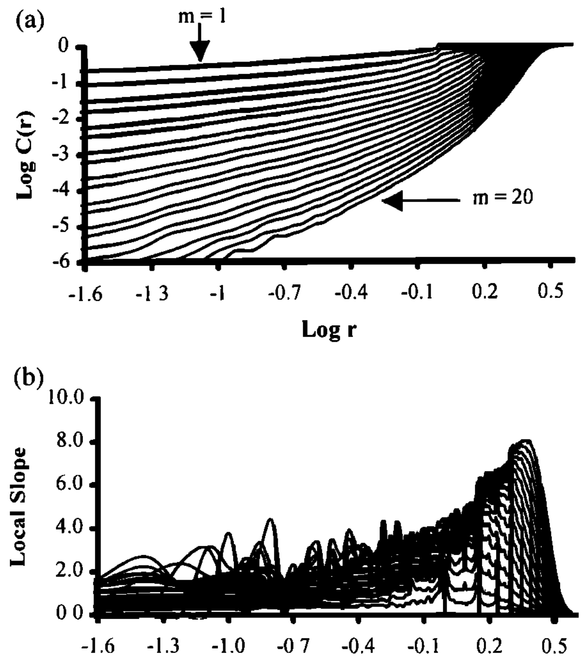

Figure 3. (a) Log $C(r)$ versus log r plot for distributions of weights of rainfall data between 12-hour and 6-hour resolutions. (b) Local slope versus log r plot for distributions of weights of rainfall data between 12-hour and 6-hour resolutions.

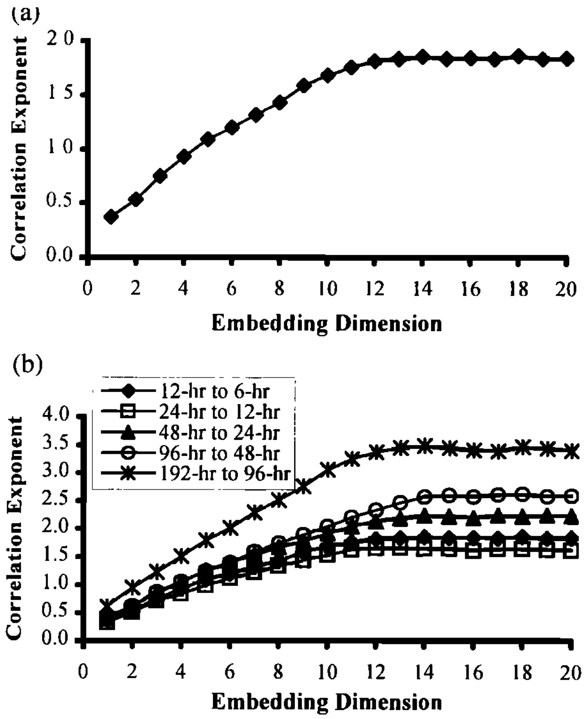

Figure 4. (a) Relationship between correlation exponent and embedding dimension for distributions of weights of rainfall data between 12-hour and 6-hour resolutions. (b) Relationship between correlation exponent and embedding dimension for distributions of weights of rainfall data between different successively doubled resolutions.

sion is an indication that the distributions of weights between rainfall at 12-hour and 6-hour resolutions exhibit low-dimensional chaotic behavior. Figure 4b shows the relationship between the correlation exponent values and the embedding dimension values for the distributions of weights between rainfall data observed at the different successively doubled resolutions considered in the present study. Saturation of the correlation exponents is observed in all the cases, indicating that the distributions of weights between rainfall data observed at different resolutions exhibit low-dimensional chaotic behavior. The correlation dimension results achieved for the different sets of distributions of weights are presented in Table 3. The existence of low-dimensional chaotic behavior in the transformation of data between different successive resolutions analyzed seems to suggest the applicability (or suitability) of a chaotic approach for rainfall disaggregation.

Though the analysis of distributions of weights between data of different successive resolutions yielded low correlation dimension estimates, variations in the dimensions are observed, as can be seen in Table 3. The highest correlation dimension value of about 3.46 is obtained for distributions of weights between rainfall data observed at 192-hour and 96-hour resolutions. The dimension decreases for resolutions between 96 and 48 hours and 48 and 24 hours and finally reaches the lowest value of about 1.65 between 24 and 12 hours. However, the same trend is not observed for the distributions of weights between 12-hour and 6-hour resolutions, instead an increase in the dimension is observed. A further increase in the dimension may be expected for distributions of weights of data between still higher and higher resolutions (e.g., between 6 and 3 hours, 3 and 1.5 hours, and so on). Unfortunately, this could not be verified in the present study, since the highest-resolution data available are only at 6-hour intervals.

The above observations on the variations of correlation dimension estimates deserve some discussion here. The (correlation) dimension of a data set is a representation of the variability or irregularity of the values in the data set that is investigated. A data set with high variability in the values provides a high dimension, which, in turn, indicates the higher complexity of the dynamics of the phenomenon. A low dimension would be the result of low variability, indicating less complexity of the dynamics of the phenomenon. The dimension results achieved in the present study indicate that the distributions of weights between rainfall data observed at 192-hour and 96-hour resolutions have a higher degree of variability than those observed at other resolutions. The correlation dimension value of about 3.46 obtained indicates that the minimum number of variables necessary to model the dynamics of the distributions of weights is 4. However, the minimum variability seems to occur for distributions of weights between rainfall data observed at 24-hour and 12-hour resolutions. In this case the correlation dimension value is about 1.65, and therefore the minimum number of variables necessary is 2. For data of the other resolutions considered in the present study, the minimum number of variables necessary to model the dynamics of the distributions of weights is 2 or 3 (Table 3).

The implications of the dimension results achieved in the present study may be the following: (1) There is only a particular range of scales or resolutions (in this case between 24 hours and 12 hours) at which the disaggregation procedure is more effective than at any other range. In other words, it may not be possible to achieve realistic results when one attempts disaggregation using very low-resolution data (e.g., 192 hours). Therefore the identification of this particular range of scales that can provide reasonably good results is important to obtain guidelines on the possibility of disaggregation. The correlation dimension could be an important tool for such identification, as this can provide important information about the complexity of the dynamics of transformation of data from one scale to another. (2) The random cascade schemes, which treat the data as a realization of a stochastic process and also assume that statistical properties of the process observed at different scales or resolutions are governed by the same relationship, may or may not be valid. The properties of the process at

Table 3. Results of Correlation Dimension Analysis of Weights

Parameter	192-hour to 96-hour Resolution	96-hour to 48-hour Resolution	48-hour to 24-hour Resolution	24-hour to 12-hour Resolution	12-hour to 6-hour Resolution
Delay time	1	1	1	1	1
Correlation dimension	3.46	2.61	2.23	1.65	1.86
Number of variables	4	3	3	2	2

different scales may be governed by a chaotic relationship rather than by a stochastic relationship, with the dimension estimate indicating the simplicity (or complexity) of the relationship and the number of variables essential to model the dynamics of transformation.

The above results for the weights of rainfall data between the different resolutions also support the earlier argument that the presence of reasonably small percentage of zeros in the time series would result in only a slight underestimation of the dimension, if any, and is explained as follows. The weights between data of 48-hour and 24-hour resolutions consist of 19.91% of zeros, and the percentage of zeros in the weights between data of 96-hour and 48-hour resolutions is 24.72%. If the percentage of zeros alone is considered responsible for the underestimation of the dimension, then the former series should yield (at least) a slightly higher dimension than the latter series. However, the dimension obtained for the former series is about 2.23, which is lower than the dimension of about 2.61 obtained for the latter series. Such an observation indicates that the dimensions of weights are not heavily underestimated because of the presence of zeros, though there may be a possibility of a slight underestimation. However, even if such an underestimation occurs, then it seems to occur almost equally for all the series as the percentage of zeros in the five different series lie in a narrow range between 14.97% and 23.17%. However, if the zero values are also considered in obtaining the distributions of weights, then the percentage of zeros in the different series might lie in a wide range between 20% and 75%, and the underestimation of the dimension could be significantly different for different time series. Such an observation is made by Sivakumar (submitted manuscript, 2000), while investigating data of four different resolutions observed in Leaf River basin. All these results suggest that it would be desirable to exclude the zero values (particularly when large in number) to avoid underestimation of the dimension, if and only if their exclusion does not result in serious consequences. As explained previously, since the presence of zero rainfall values of a particular resolution do not contribute anything to the disaggregation to another (higher) resolution but only causes an underestimation of the dimension and problems in disaggregation procedure, their exclusion seems not only reasonable but also necessary.

3. Chaotic Disaggregation Model for Rainfall

3.1. Model Formulation

It should be clear, from the correlation dimension results presented in section 2, that the dynamics of transformation of rainfall data between different resolutions may be better described by a chaotic approach rather than a stochastic approach. In this spirit an attempt is made in this section to develop a disaggregation model for rainfall based on a chaotic dynamical approach. The effectiveness of the model is tested by employing it to analyze the rainfall data of different successively doubled resolutions, mentioned in section 2.3.3.

Let us assume that we have a rainfall time series X_i, $i = 1, 2, \ldots, N$, at a certain resolution T_1, and the task at hand is to obtain the rainfall values $(Z_i)_k$, $k = 1, 2, \ldots, p$, at a higher resolution T_2, where $p = T_1/T_2$. Let us also assume that the values of X_i are distributed into $(Z_i)_k$ according to $(Z_i)_k = (W_i)_k X_i$, where $(W_i)_k$ are the distributions of weights of X_i to $(Z_i)_k$ and $\Sigma_{k=1}^{p} (W_i)_k = 1$. Assuming also that information is available about the history of distributions of weights $(W_i)_k$ (or X_i and $(Z_i)_k$), $i = 1, 2, \ldots, n$, where $n < N$, an attempt is made here, based on the ideas gained from chaotic dynamical theory, to formulate a model to obtain the future distributions of weights $(W_i)_k$ and hence the values of $(Z_i)_k$, $i = n + 1, \ldots, N$ (to make the parameters involved in the model consistent with the ones used in the correlation dimension method, the total number of points is always kept equal to N; the history of n rainfall values (training set), where $n < N$, is used to disaggregate the rainfall values from $n + 1$ to N). The approach adopted in the formulation of the chaotic disaggregation model is somewhat similar to the one adopted in the nonlinear prediction method [e.g., *Farmer and Sidorowich*, 1987; *Casdagli*, 1989, 1991; *Sugihara and May*, 1990] and is explained as follows.

Let us consider determining first the distributions of weights $(W_{n+1})_k$. The first step is to reconstruct the phase space of the time series X_i, $i = 1, 2, \ldots, n + 1$, observed at the resolution T_1, according to

$$\mathbf{Y}_j = (X_j, X_{j+\tau}, X_{j+2\tau}, \ldots, X_{j+(m-1)\tau}), \tag{5}$$

where $j = 1, 2, \ldots, (n + 1) - (m - 1)\tau/\Delta t$. The next step is to assume a functional relationship between the vectors $\mathbf{Y}_j$, such as

$$\mathbf{Y}_{j+T} = F_T(\mathbf{Y}_j). \tag{6}$$

The problem then is to find an appropriate expression for F_T. There are several possible approaches for determining F_T, but the most promising seems to be the "local approximation method" [e.g., *Farmer and Sidorowich*, 1987; *Casdagli*, 1989; *Sugihara and May*, 1990], which uses only nearby states. The basic idea in the local approximation method is to break up the domain into local neighborhoods and fit parameters in each neighborhood separately. The disaggregation of X_{n+1} is made based on $\mathbf{Y}_j$, $j = (n + 1) - (m - 1)\tau/\Delta t$, and its neighbors $\mathbf{Y}'_j$ for all $j' < j$. The neighbors of $\mathbf{Y}_j$ are found on the basis of the minimum values of $\|\mathbf{Y}_j - \mathbf{Y}'_j\|$. If only one neighbor is considered, then the distributions of weights $(W_{n+1})_k$ of X_{n+1} would be the distributions of weights of the corresponding element X_j in the nearest vector Y'_j. This is called the zeroth order or nearest-neighbor approximation. However, such an approach might not give good results, since a single neighbor might not involve the important dynamics of the system. An improvement to the zeroth-order approximation is the first-order approximation. This approach considers k' number of neighbors, and the distributions of weights $(W_{n+1})_k$ of X_{n+1} is taken as an average of the k' values' distributions of weights of the corresponding elements X_j in the nearest vectors. The optimal value of k' (k'_{opt}) is generally determined by trial and error. Having determined the distributions of weights, the disaggregation of the rainfall value X_{n+1} observed at the resolution T_1 to rainfall values $(Z_{n+1})_k$ at resolution T_2 can be obtained according to $(Z_{n+1})_k = (W_{n+1})_k X_{n+1}$. The above procedure can be repeated to obtain the distributions of weights of rainfall values X_{n+2}, X_{n+3}, $\ldots$, X_N, that is, $(W_{n+2})_k$, $(W_{n+3})_k$, $\ldots$, $(W_N)_k$, and hence the rainfall values at the resolution T_2, that is, $(Z_{n+2})_k$, $(Z_{n+3})_k$, $\ldots$, $(Z_N)_k$. The accuracy of the disaggregation can be evaluated by using any of the standard statistical measures. In this study, the linear correlation coefficient between the modeled values and the observed values is used as a main tool to determine the accuracy of disaggregation. A correlation coefficient of 1 (one) is considered as a perfect disaggregation, whereas a value of 0 (zero) refers to no relationship between the modeled and the observed values. In addition to the linear correlation coefficient, root-mean-square error, coefficient of efficiency, time series plots, and scatter diagrams are also used to choose the best disaggregation results, among a large combination of results achieved with varying number of neighbors and embedding dimensions.

3.2. Identification of Chaos (Inverse Approach)

The disaggregation model formulated above also has the potential advantage in that the disaggregation results themselves can be used to identify the behavior (chaotic or stochastic) of dynamics of rainfall transformation from one scale to another. The identification can be done in two possible ways by checking the accuracy of disaggregation (i.e., correlation coefficient or root-mean-square error or coefficient of efficiency) against (1) the number of neighbors and (2) the embedding dimension. The identification of chaos using the disaggregation results themselves can therefore be termed as an inverse approach. The inverse approach used in the present study to identify chaos in rainfall transformation is similar to the one proposed by *Casdagli* [1989, 1991] and used by, for example, *Sivakumar et al.* [1999a], except that the present study uses the accuracy of disaggregation instead of prediction, used in those studies.

The concept behind the use of number of neighbors is that a smaller number of neighbors corresponds to a deterministic modeling approach, whereas a larger number of neighbors corresponds to a stochastic modeling approach [e.g., *Casdagli*, 1991]. Therefore if models near to the deterministic extreme (i.e., smaller number of neighbors) provide more accurate disaggregation results compared to those near the stochastic extreme (i.e., larger number of neighbors), then this may be considered as an evidence of low-dimensional chaos in the data. If models near the stochastic extreme provide more accurate disaggregation results compared to those near to the deterministic extreme, then this may be considered as an indication of stochasticity.

With respect to the embedding dimension the presence (or absence) of chaos can be identified as follows. If the dynamics is chaotic, then it is expected that the disaggregation accuracy would increase (to reach its best) with the increase in the embedding dimension up to a certain point, called the optimal embedding dimension (m_{opt}), and remain close to its best for embedding dimensions higher than m_{opt}. However, for stochastic time series, there would be no increase in the disaggregation accuracy with an increase in the embedding dimension, and the accuracy would remain the same for any value of the embedding dimension [e.g., *Casdagli*, 1989].

3.3. Model Application to Rainfall Data

The chaotic disaggregation model formulated above is employed to analyze rainfall data of successively doubled resolutions (i.e., $p = 2$) observed at Leaf River basin (see sections 2.3.2 and 2.3.3 for details). Again, only the nonzero rainfall values are considered in the disaggregation procedure. In the present study, the number of neighbors considered in the disaggregation model ranges from 1 to 200, and the procedure is carried out for embedding dimension values from 1 to 10. In each case (i.e., disaggregation between different successively doubled resolutions) the (lower resolution) time series is split into two parts, with the last 100 points that are disaggregated to yield 200 (higher resolution) values, forming the test set to

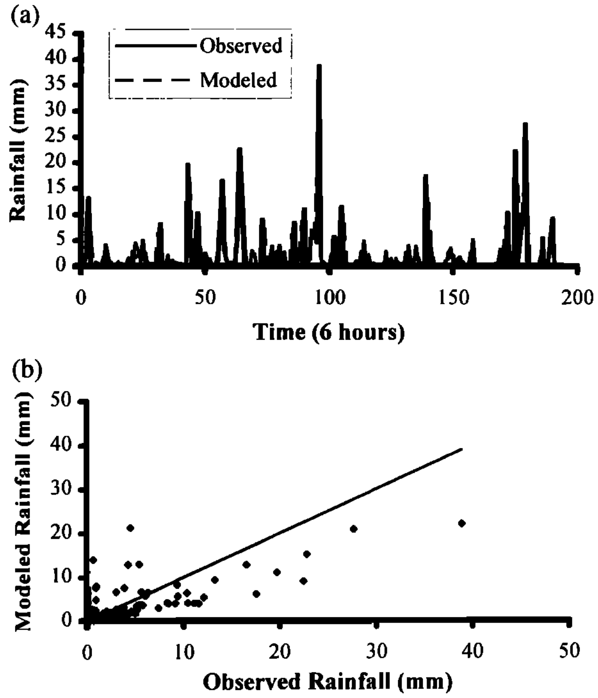

Figure 5. (a) Comparison between time series plots of modeled and observed rainfall disaggregation values from 12-hour resolution to 6-hour resolution. (b) Scatterplot of modeled versus observed rainfall disaggregation values from 12-hour resolution to 6-hour resolution.

evaluate the disaggregation accuracy. The discussion here revolves around two important aspects: (1) effectiveness (or suitability) of the chaotic disaggregation model for rainfall and (2) identification of chaos from the disaggregation results, using the inverse approach explained in section 3.2.

3.3.1. Model effectiveness. Figure 5a, for example, presents a comparison of the time series between the modeled values and the observed values when the rainfall data observed at 12-hour resolution is disaggregated into 6-hour resolution. The scatterplot of the modeled and the observed values is presented in Figure 5b, with the solid 1:1 (diagonal) line for reference. The disaggregation results shown in Figures 5a and 5b are those obtained for the optimal number of neighbors ($k'_{\text{opt}} = 10$) and embedding dimension ($m_{\text{opt}} = 4$) (see Table 4). The correlation coefficient value achieved is about 0.723, and the root-mean-square error and coefficient of efficiency values are respectively 3.42 and 0.567 (see Table 4). Figures 5a and 5b show fairly accurate matching between the modeled and the observed values, except when the observed rainfall is very high or zero. In the latter case a comparison between the modeled and the observed rainfall values indicates that the modeled rainfall has significant number of nonzero values against the zero observed values. (Similar results are achieved also for data of other resolutions.) A possible explanation for this (overestimation) may be the following. The zero rainfall values observed at the 6-hour resolution are the result of distributions of nonzero rainfall values observed at the 12-hour resolution, which means that the distributions of weights are 1 and 0 (or 0 and 1) (see Figure 2). However, in the disaggregation model used in the present study, the distributions of weights for a rainfall value are obtained by averaging those of the rainfall values nearest to the reconstructed vector consist-

Table 4. Results of Rainfall Disaggregation

Parameter	192-hour to 96-hour Resolution	96-hour to 48-hour Resolution	48-hour to 24-hour Resolution	24-hour to 12-hour Resolution	12-hour to 6-hour Resolution
Correlation dimension	3.46	2.61	2.23	1.65	1.86
Correlation coefficient	0.7148	0.6772	0.7165	0.7530	0.7234
Root-mean-square error	15.62	14.41	9.91	4.89	3.42
Coefficient of efficiency	0.401	0.458	0.507	0.588	0.567
Optimal dimension	6	6	4	4	4
Optimal neighbors	10	10	10	10	10

ing of that particular rainfall value, as explained in section 3.1. The consideration of several nearest rainfall values could result in distributions of weights other than 1 and 0 (or 0 and 1), if those nearest values have distributions of weights other than 1 and 0 (or 0 and 1). A similar explanation can be provided also for the bias (underestimation) achieved when the observed values are high (or even moderate). In such a case the distributions of weights could be highly uneven, that is, closer to 1 and 0 (or 0 and 1), since an averaging procedure might bring the weights around the mean of 1 and 0. Another possible argument for this underestimation may be that it could also be the result of insufficient number of neighbors, since the consideration of only a small number of neighbors may not be sufficient to bring the weights significantly different than the mean of 1 and 0. However, the results achieved from the present study for data of different resolutions analyzed indicate that the optimal results are achieved only when the number of neighbors is low (i.e., $k' \sim 10$) (see section 3.3.2).

Table 4 presents a summary of the disaggregation results achieved, using the chaotic model, for data of different successively doubled resolutions considered in the present study. The correlation coefficient, root-mean-square error, and coefficient of efficiency values presented in Table 4 are those achieved for the optimal number of neighbors and embedding dimensions, also presented therein. In most of the cases, for example, the correlation coefficients between the modeled and the observed rainfall values are above 0.70, indicating the effectiveness of the chaotic disaggregation procedure for rainfall. However, the variations observed in the correlation coefficients when disaggregation is between different successively doubled resolutions seem to be interesting. As Table 4 reveals, the highest correlation coefficient of about 0.753 is achieved when the disaggregation is from 24-hour resolution to 12-hour resolution, whereas the lowest coefficient of about 0.677 occurs when the disaggregation is from 96-hour resolution to 48-hour resolution. Also, there seems to be a trend of a decrease in the correlation coefficient when the disaggregation resolutions are higher than 24 hours (i.e., between 48 and 24 hours and 96 and 48 hours), except between 192 hours and 96 hours. Such a trend is also supported by the coefficient of efficiency values. An insight into the disaggregation accuracy results and the correlation dimension results of the distributions of weights, also presented in Table 4, seems to indicate a possible inverse relationship between the two. The best disaggregation accuracy (e.g., highest correlation coefficient) achieved for disaggregation from 24 to 12 hours corresponds to the lowest correlation dimension achieved for the distributions of weights (about 1.65). The trend seems to be present throughout the resolutions of data considered for analysis, except for those between 192 and 96 hours. The inverse relationship between the correlation dimension and the correlation coefficient is consistent with the concept of the correlation dimension method presented earlier. That is, a higher dimension is an indication of a more complex process (or transformation dynamics), which is more difficult to model than a less complex process, which is recognized by a lower dimension. All these observations suggest the usefulness and validity of the concept of correlation dimension, which is developed in the context of chaos theory, for understanding the behavior of the dynamics involved in rainfall transformation and its subsequent modeling.

3.3.2. Evidence of chaos (inverse approach). The above results regarding the existence of chaos in the dynamics of rainfall transformation can be supported further using the inverse approach. Figure 6a shows, for example, the relationship between the correlation coefficient and the number of neighbors when the rainfall data at 12-hour resolution is disaggregated into 6-hour resolution. The disaggregation results shown are those obtained for the optimal embedding dimension value ($m_{\text{opt}} = 4$). Figure 6a shows that the correlation coefficient increases with the increase in the number of neighbors up to a

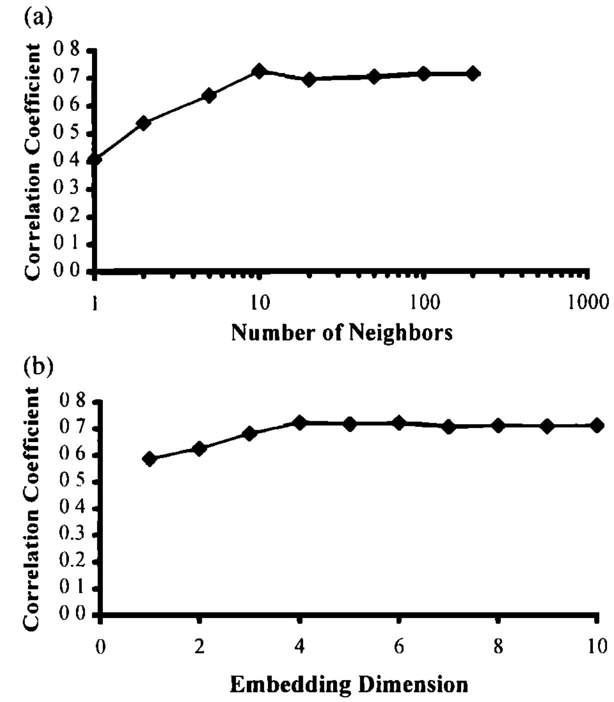

Figure 6. (a) Relationship between correlation coefficient and number of neighbors: Disaggregation of rainfall data from 12-hour resolution to 6-hour resolution. (b) Relationship between correlation coefficient and embedding dimension: Disaggregation of rainfall data from 12-hour resolution to 6-hour resolution.

certain point ($k' = 10$) and then seems to remain constant when the number of neighbors is increased further. Though the correlation coefficient values seem to remain the same for number of neighbors equal to 10 or more, the best results seem to occur only when the number of neighbors is equal to 10 (as observed from the time series plots and scatterplots). The observation of best disaggregation results obtained with a smaller number of neighbors (i.e., a deterministic extreme) may be considered to provide evidence of chaos in the transformation dynamics of rainfall data between 12-hour and 6-hour resolutions. Similar results are achieved also for data of other resolutions considered in the present study, suggesting the existence of chaos.

Figure 6b shows the variation of the correlation coefficient against the embedding dimension for disaggregation of rainfall from 12-hour resolution to 6-hour resolution, with the number of neighbors equal to the optimal number of neighbors obtained above ($k'_{opt} = 10$). The correlation coefficient increases with the increase in the embedding dimension up to a certain value ($m = 4$) and then remains saturated beyond that value. Similar results are also observed for the coefficient of efficiency values. Again, the best results are obtained when the embedding dimension is 4 (i.e., $m_{opt} = 4$). The disaggregation results achieved for rainfall data of the other different resolutions considered in the present study also indicate optimal embedding dimensions (see Table 4), providing evidence of the existence of chaotic behavior in the dynamics of rainfall transformation. The optimal embedding dimensions thus obtained seem to be in good agreement with the one suggested by *Takens* [1980], that is, $2d + 1$, where d is the dimension of the time series. All these results only provide further support to the results achieved using the correlation dimension method regarding the existence of chaos in the dynamics of the transformation of rainfall data between successively doubled resolutions observed in Leaf River basin. Also, the good agreement between the optimal embedding dimensions and the correlation dimensions seems to suggest that the estimated correlation dimensions could well be closer to the actual dimension of rainfall transformation process and that there is only a slight underestimation of the dimension, if any.

4. Summary and Conclusions

An important limitation often faced in the investigation of hydrological, meteorological, and agricultural processes is the lack of high-resolution data to understand the intricacies of the processes and describe them in a sophisticated and physically accurate way. Though the past decade has witnessed a number of studies to solve this problem, the main concern of such studies seems to be the inability of the stochastic approaches used therein to establish connections between the structure of the models and the underlying dynamics of the processes [*Olsson*, 1998]. The present study was aimed at the possibility of establishing such connections through the introduction of a chaotic dynamical framework. The chaotic dynamical framework introduced was based on the notion that seemingly irregular-looking processes could be the result of simple deterministic systems influenced by a few nonlinear interdependent variables. Rainfall data of six different successively doubled resolutions observed at Leaf River basin in the state of Mississippi, United States of America, were considered for the investigation. Specifically, the distributions of weights between any two successively doubled resolutions were studied. The correlation dimension method was used to investigate the behavior of the dynamics of rainfall transformation.

The analysis of the distributions of weights of transformation of rainfall data between successively doubled resolutions yielded finite and low correlation dimensions and thus provided preliminary evidence on the existence of chaotic behavior in the transformation process and therefore suggested the applicability (or suitability) of a chaotic dynamical approach. The correlation dimension results also provided important information regarding the identification of a particular range of resolutions at which rainfall disaggregation could be very effective. With such encouraging results a chaotic disaggregation model for rainfall was then formulated, based on an approach that is somewhat similar to the one generally used for the prediction of chaotic time series, and then employed to analyze the Leaf River rainfall data. The inverse approach, where the disaggregation results themselves can be used to identify chaos, indicated the existence of chaotic behavior in the rainfall transformation, thus providing additional support for the results obtained using the correlation dimension method. The correlation coefficients (i.e., disaggregation accuracy) between the modeled and the observed values were found to be about 0.70 (and also reasonably low root-mean-square error and high coefficient of efficiency values), indicating the effectiveness of the chaotic approach for rainfall disaggregation. It was also found, however, that the model was not able to simulate well the very high and zero rainfall values. This could have possibly been due to the averaging procedure employed in the model formulation to identify the distributions of weights of the nearest neighbors in the reconstructed phase space, and therefore it is believed that the results can be significantly improved by employing a better approach. However, the presence of noise in the rainfall data, such as the measurement error, could have also had significant influence on the outcomes of the disaggregation model, and therefore a reduction/removal of noise from the data could be useful to improve the disaggregation accuracy.

The study of the rainfall transformation between data of different successively doubled resolutions observed at Leaf River basin provided evidence regarding the existence of the chaotic nature of the dynamics. The minimum number of variables essential to model the dynamics of the transformations was in the range between 2 and 4, and the optimal number of variables was identified to lie between 4 and 6. These results suggest that the seemingly irregular behavior of the distributions of weights between data of different resolutions could, indeed, be the result of a simple deterministic system with a few nonlinear interdependent variables. The results seem to suggest that a chaotic framework could be more suitable than a stochastic representation for modeling the dynamics of rainfall transformation. This study, however, was, to the authors' knowledge, the first ever study conducted on the investigation of rainfall transformation based on the ideas gained from the theory of chaos. The authors realize, however, that because of the limitations of the correlation dimension method [e.g., *Osborne and Provenzale*, 1989] (or any other chaos identification method for that matter), it would be necessary to employ additional (and diverse) techniques to substantiate the results achieved thus far regarding the existence of chaos in rainfall transformation. However, though more emphasis was given in the present study to the investigation of the presence of chaos in rainfall transformation than its modeling, it is interesting that the application of even a very simple chaotic model yielded reasonable disaggregation results. Such outcomes are very encouraging, and therefore the concept of deterministic chaos seems to have great practical potential in rainfall transformation. Investigations using other techniques to support the present results regarding the presence of chaos in rainfall transformation, and efforts to improve the disaggregation model formulated in the present study, are underway, details of which will be discussed elsewhere.

Acknowledgments. The work presented in this paper was carried out with the financial support of ICSC World Laboratory, Switzerland, in cooperation with the Department of Hydrology and Water Resources, University of Arizona, and support of NASA grant NAGW2425. The first author wishes to thank the World Laboratory and Soroosh Sorooshian for the Harshbarger Fellowship awarded to him. The useful suggestions given by Ronny Berndtsson and an anonymous reviewer are gratefully acknowledged.

References

Abarbanel, H. D. I., R. Brown, and J. B. Kadtke, Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra, *Phys. Rev. A*, *41*, 1782–1807, 1990.

Arnold, J. G., and J. R. Williams, Stochastic generation of internal storm structure, *Trans. ASAE*, *32*(1), 161–166, 1989.

Berndtsson, R., K. Jinno, A. Kawamura, J. Olsson, and S. Xu, Dynamical systems theory applied to long-term temperature and precipitation time series, *Trends Hydrol.*, *1*, 291–297, 1994.

Bo, Z., S. Islam, and E. A. B. Eltahir, Aggregation-disaggregation properties of a stochastic rainfall model, *Water Resour. Res.*, *30*(12), 3423–3435, 1994.

Carsteanu, A., and E. Foufoula-Georgiou, Assessing dependence of weights in a multiplicative cascade model of temporal rainfall, *J. Geophys. Res.*, *101*, 26,363–26,370, 1996.

Casdagli, M., Nonlinear prediction of chaotic time series, *Physica D*, *35*, 335–356, 1989.

Casdagli, M., Chaos and deterministic versus stochastic non-linear modeling, *J. R. Stat. Soc., Ser. B*, *54*(2), 303–328, 1991.

Connolly, R. D., J. Schirmer, and P. K. Dunn, A daily rainfall disaggregation model, *Agric. For. Meteorol.*, *92*, 105–117, 1998.

Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley, Stochastic point process modeling of rainfall, II, Regionalization and disaggregation, *J. Hydrol.*, *175*, 47–65, 1996.

Econopouly, T. W., D. R. Davis, and D. A. Woolhiser, Parameter transferability for a daily rainfall disaggregation model, *J. Hydrol.*, *118*, 209–228, 1990.

Farmer, D. J., and J. J. Sidorowich, Predicting chaotic time series, *Phys. Rev. Lett.*, *59*, 845–848, 1987.

Fraedrich, K., Estimating the dimensions of weather and climate attractors, *J. Atmos. Sci.*, *43*(5), 419–432, 1986.

Glasbey, C. A., G. Cooper, and M. B. McGechan, Disaggregation of daily rainfall by conditional simulation from a point-process model, *J. Hydrol.*, *165*, 1–9, 1995.

Grassberger, P., and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D*, *9*, 189–208, 1983a.

Grassberger, P., and I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.*, *50*, 346–349, 1983b.

Grassberger, P., and I. Procaccia, Estimation of the Kolmogorov entropy from a chaotic signal, *Phys. Rev. A*, *28*, 2591–2593, 1983c.

Gupta, V. K., and E. Waymire, A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–267, 1993.

Hense, A., On the possible existence of a strange attractor for the southern oscillation, *Beitr. Phys. Atmos.*, *60*(1), 34–47, 1987.

Hershenhorn, J., and D. A. Woolhiser, Disaggregation of daily rainfall, *J. Hydrol.*, *95*, 299–322, 1987.

Holzfuss, J., and G. Mayer-Kress, An approach to error-estimation in the application of dimension algorithms, in *Dimensions and Entropies in Chaotic Systems*, edited by G. Mayer-Kress, pp. 114–122, Springer-Verlag, New York, 1986.

Jayawardena, A. W., and F. Lai, Analysis and prediction of chaos in rainfall and stream flow time series, *J. Hydrol.*, *153*, 23–52, 1994.

Mandelbrot, B. B., Intermittent turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier, *J. Fluid Mech.*, *62*, 331–358, 1974.

Olsson, J., Evaluation of a scaling cascade model for temporal rainfall disaggregation, *Hydrol. Earth Syst. Sci.*, *2*(1), 19–30, 1998.

Olsson, J., and R. Berndtsson, Temporal rainfall disaggregation based on scaling properties, *Water Sci. Technol.*, *37*(11), 73–79, 1998.

Osborne, A. R., and A. Provenzale, Finite correlation dimension for stochastic systems with power-law spectra, *Physica D*, *35*, 357–381, 1989.

Packard, N. H., J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a time series, *Phys. Rev. Lett.*, *45*, 712–716, 1980.

Palus, M., Testing for nonlinearity using redundancies: Quantitative and qualitative aspects, *Physica D*, *80*, 186–205, 1995.

Perica, S., and E. Foufoula-Georgiou, Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions, *J. Geophys. Res.*, *101*, 26,347–26,361, 1996.

Porporato, A., and L. Ridolfi, Clues to the existence of deterministic chaos in river flow, *Int. J. Mod. Phys. B*, *10*, 1821–1862, 1996.

Porporato, A., and L. Ridolfi, Nonlinear analysis of river flow time sequences, *Water Resour. Res.*, *33*(6), 1353–1367, 1997.

Prichard, D., and J. Theiler, Generalized redundancies for time series analysis, *Physica D*, *84*, 476–493, 1995.

Puente, C. E., and N. Obregon, A deterministic geometric representation of temporal rainfall: Results for a storm in Boston, *Water Resour. Res.*, *32*(9), 2825–2839, 1996.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham, Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London, Ser. A*, *410*, 269–288, 1987.

Rodriguez-Iturbe, I., D. R. Cox, and V. Isham, A point process model for rainfall: Further developments, *Proc. R. Soc. London, Ser. A*, *417*, 283–298, 1988.

Rodriguez-Iturbe, I., F. B. De Power, M. B. Sharifi, and K. P. Georgakakos, Chaos in rainfall, *Water Resour. Res.*, *25*(7), 1667–1675, 1989.

Sangoyomi, T. B., U. Lall, and H. D. I. Abarbanel, Nonlinear dynamics of the Great Salt Lake: Dimension estimation, *Water Resour. Res.*, *32*(1), 149–159, 1996.

Schertzer, D., and S. Lovejoy, Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, *J. Geophys. Res.*, *92*, 9693–9714, 1987.

Sharifi, M. B., K. P. Georgakakos, and I. Rodriguez-Iturbe, Evidence of deterministic chaos in the pulse of storm rainfall, *J. Atmos. Sci.*, *47*, 888–893, 1990.

Sivakumar, B., Identification of chaos and influence of noise on prediction: Singapore rainfall, Ph.D. thesis, Dep. of Civ. Eng., Natl. Univ. of Singapore, Singapore, 1999.

Sivakumar, B., Chaos theory in hydrology: Important issues and interpretations, *J. Hydrol.*, *227*, 1–20, 2000.

Sivakumar, B., S. Y. Liong, and C. Y. Liaw, Evidence of chaotic behavior in Singapore rainfall, *J. Am. Water Resour. Assoc.*, *34*(2), 301–310, 1998.

Sivakumar, B., S. Y. Liong, C. Y. Liaw, and K. K. Phoon, Singapore rainfall behavior: Chaotic?, *J. Hydrol. Eng.*, *4*(1), 38–48, 1999a.

Sivakumar, B., K. K. Phoon, S. Y. Liong, and C. Y. Liaw, A systematic approach to noise reduction in chaotic hydrological time series, *J. Hydrol.*, *219*, 103–135, 1999b.

Sugihara, G., and R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, *344*, 734–741, 1990.

Takens, F., Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics 898*, edited by D. A. Rand and L. S. Young, pp. 366–381, Springer-Verlag, New York, 1980.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, Testing for nonlinearity in time series: The method of surrogate data, *Physica D*, *58*, 77–94, 1992.

Tsonis, A. A., J. B. Elsner, and K. P. Georgakakos, Estimating the dimension of weather and climate attractors: Important issues about the procedure and interpretation, *J. Atmos. Sci.*, *50*, 2549–2555, 1993.

Wolf, A., J. B. Swift, H. L. Swinney, and A. Vastano, Determining Lyapunov exponents from a time series, *Physica D*, *16*, 285–317, 1985.

X. Gao, H. V. Gupta, and S. Sorooshian, Department of Hydrology and Water Resources, University of Arizona, Box 210011, Tucson, AZ 85721-0011. (gao@hwr.arizona.edu; hoshin@hwr.arizona.edu; soroosh@hwr.arizona.edu)

B. Sivakumar, Department of Land, Air and Water Resources, University of California, Veihmeyer Hall, Davis, CA 95616. (sbellie@ucdavis.edu)

(Received May 10, 2000; revised June 21, 2000; accepted June 27, 2000.)